AGE-SPECIFIC ADJUSTMENT OF GRADUATED MORTALITY

BY

YAHIA SALHI AND PIERRE-E. THÉROND

ABSTRACT

Recently, there has been an increasing interest from life insurers to assess their portfolios' mortality risks. The new European prudential regulation, namely Solvency II, emphasized the need to use mortality and life tables that best capture and reflect the experienced mortality, and thus policyholders' actual risk profiles, in order to adequately quantify the underlying risk. Therefore, building a mortality table based on the experience of the portfolio is highly recommended and, for this purpose, various approaches have been introduced into actuarial literature. Although such approaches succeed in capturing the main features, it remains difficult to assess the mortality when the underlying portfolio lacks sufficient exposure. In this paper, we propose graduating the mortality curve using an adaptive procedure based on the local likelihood. The latter has the ability to model the mortality patterns even in presence of complex structures and avoids relying on expert opinions. However, such a technique fails to offer a consistent yet regular structure for portfolios with limited deaths. Although the technique borrows the information from the adjacent ages, it is sometimes not sufficient to produce a robust life table. In the presence of such a bias, we propose adjusting the corresponding curve, at the age level, based on a credibility approach. This consists in reviewing the assumption of the mortality curve as new observations arrive. We derive the updating procedure and investigate its benefits of using the latter instead of a sole graduation based on real datasets. Moreover, we look at the divergences in the mortality forecasts generated by the classic credibility approaches including Hardy–Panjer, the Poisson–Gamma model and the Makeham framework on portfolios originating from various French insurance companies.

KEYWORDS

Credibility, mortality, life insurance, graduation, smoothing, local likelihood, prediction.

1. INTRODUCTION

For insurers, the assessment of experienced mortality is of paramount importance. The new regulations and norms, established under Solvency II, shed light

Astin Bulletin 48(2), 543-569. doi: 10.1017/asb.2018.4

on the need of life tables that best reflect the experience of insured portfolios own mortality in order to reliably quantify the underlying risk. Insurers, in France for example, are used to rely on regulatory life tables for pricing and reserving purposes, which are sometimes too conservative. However, the use of inadequate life tables, i.e. that are too conservative, would considerably affect the financial profitability of life insurance businesses as well as insurers' competitiveness. Among others, from a Solvency II perspective, using overly conservative tables lead to two major impacts: (i) an increase of *best estimate* technical provisions (and thus a decrease of *Basic Own-Funds*), and (ii) an increase of the *base figure* used for calculating the capital charge for mortality risk.

A natural and straightforward approach to handle this issue is to use the available data at the portfolio level and build an entity-specific mortality table. However, practitioners may face technical difficulties related to the size of the portfolios and the heterogeneity of the guarantees (for the same underlying risk, say mortality risk). For instance, an insurer may have a fairly big portfolio but with policyholders holding different insurance contracts: pure endowment contracts, unit-linked contracts with minimum death guarantees, loan insurance and so on. In such a case, it is difficult to build a mortality table based on the sole experience of each product or guarantee. More precisely, the constructed table would not be able to represent the mortality profile of the policyholders thus failing to capture the underlying risk. This should also be the case even if the mortality table is periodically updated with the incoming new data. If a mortality table only based on the experience stemming from one product or guarantee is drawn, then there will be a sample size problem. The latter arises not only at the portfolio level but also for individual ages. In fact, the mortality profile is highly dependent on the age of the individuals and some age groups being *poorly* represented may alter the quantification of the mortality risk at each individual age.

In this paper, we consider an insurer with exposures to different coverages and aiming to establish an experience-based mortality table for each policy and age level, as individuals may have different risk profiles (as showed by some empirical mortality studies, e.g. see Vaupel *et al.* (1979) and Hougaard (1984) among others). To begin with, we consider a graduation principle to build mortality rates at the insured portfolio level. There are usually two sorts of methods: non-parametric and parametric, see Forfar *et al.* (1988a) and Debón *et al.* (2006) for a comprehensive introduction to the use of both graduation techniques. The non-parametric framework is very useful in practice especially when there is sufficient data. This method relies on the use of kernel estimation techniques which were first used for graduation by Copas and Haberman (1983) and Ramlau-Hansen (1983). There is extensive literature on this subject and we may observe two schools. First, there is a continuous approach that defines data sampling via stochastic counting processes and which considers the lifetimes of individuals to be continuous random variables subject to random censorships, i.e. left truncation and right censoring. In our case, the mortality data that we use are divided into discrete yearly numbers of death occurrences and exposures. Therefore, these data only allow the use of an approach based on an approximation of the continuous filtered model. Both the continuous and the discrete formulations have been intensively explored in the literature, see e.g. Fan and Gijbels (1995), Jiang and Doksum (2003), Nielsen *et al.* (2009) and more recently Gámiz *et al.* (2016). In these models, the hazard rate is estimated using a non-parametric kernel method. A number of commonly used smoothing methods such as smoothing splines, kernel estimates and local polynomial fitting can be used to implement the basic step of the graduation of a mortality table. More recently, estimators based on local polynomial fitting, discussed in earlier works of Cleveland (1979) and Lejeune (1985), among others, have become more popular. This keen interest turned out to be useful, in particular for their good performance and analytical tractability, see for example, the monograph by Fan and Gijbels (1996).

In the approach proposed here, local polynomial fitting methods are used as an implementation of smoothing methods. This allows us to model the mortality patterns even in presence of complex structures and avoid relying on expert opinions. In Tomas (2011), the author explores the same adaptive smoothing procedure applied to the dataset used in this paper.

The graduated mortality can then be used to project future insurance liabilities related to the underlying population. However, the evolution of the flow of data related to the latest available information is not taken into account. This should be, for example, used to update the graduated mortality. Though, if one decides to redo a graduation procedure including the new data, the forecasts are likely to be unstable, adding potential volatility to the underlying reserves and capital charges. Therefore, the primary contribution of this paper is the incorporation of sample bias into the graduated mortality table model by introducing an unobserved variate for individual differences in each attained age. Such an approach has been considered in Salhi *et al.* (2016) but with different graduated curves. The latter used a parametric model, i.e. Makeham law, to first build the mortality curve and then applies a credibility procedure to a portfolio-sensitive parameter. Other approaches have also been introduced in the literature but work directly on the aggregate death counts, e.g. Hardy and Panjer (1998). Unlike the traditional approaches that focus on updating the aggregate deaths recorded within the whole portfolio, the proposed adjustment approach is intended to enhance the predictive ability of the graduated mortality using a credibility-based revision at the age level and not on the aggregate portfolio level, while borrowing information from other portfolios with sufficient information. More formally, our methodology is based on a discretization of the Nielsen and Sandqvist (2000) credibility approach. The latter, however, did not consider a multidimensional credibility method as the underlying risk does not exhibit an extra dimension rather than the observation date, see Section 3.

The rest of the paper is organized as follows. Section 2 specifies the notation and assumptions used throughout the paper. It also introduces the smoothing model in its general and continuous form. A discretization of the latter is considered as mentioned earlier and we also recall some statistical inference results used in the sequel. Section 3 introduces the credibility approach to the graduated mortality. We specify the model and illustrate the connection with recent literature. Furthermore, we derive the main tools needed to fully characterize the next prediction period of mortality rates when a (multiplicative) credibility factor is taken into account. Section 4 presents an application with experience data originating from some French insurance companies. Finally, some remarks in Section 5 conclude the paper.

2. Notation, assumptions and preliminaries

2.1. Notation, assumptions and continuous time local smoothing

Assume that we have at our disposal mortality statistics originating from K portfolios (or companies) over the time interval $[0, T_i]$, $i \in \{1, \dots, K\}$. We suppose that the portfolios are composed of I_i individuals for which we associate a triplet $(Y_e^i, Z_e^i, \Delta_e^i)$, for $e = 1, \dots, I_i$, where Y_e^i is the age that an individual enters the portfolio during the considered period, Z_e^i the age they leave the portfolio and Δ_e^i an indicator of the censoring status. In other terms, Δ_e^i is equal to 1 when the individual deceases during the period $[0, T_i]$ and 0 when they leave for other reasons, e.g. surrendering their policy. Based on this triplet, which can be observed in most life insurance portfolios, we let $N_e^i(x) = \Delta_e^i \mathbf{1}_{\{Z_e^i \leq x\}}$ be the counting process indicating the death of the individual e before age x. Similarly, we define the process $L_e^i(x) = \mathbf{1}_{\{Y_e^i \leq x < Z_e^i\}}$ that indicates if the insured is at risk at age x. For all the portfolios, we are interested in the mortality behavior in an age interval $[x_1, x_{n_i}]$. Moreover, under usual conditions, we assume Cox (1972)'s multiplicative model where the random intensity of death φ_x^i, at age x of portfolio i is related to a reference φ_x^{ref} as follows:

$$\varphi_x^i = \exp[f^i(x)]\varphi_x^{\text{ref}}, \tag{1}$$

where f^i is an unspecified, smooth and deterministic function of the age x. The latter allows us to link the mortality of the company i to the baseline at the attained age level x. Here, we adopt a parametric form for the function f^i and denote β^i this vector of parameters which will be specified later in this section.

Remark 1. *In this assumption, the baseline mortality is shared across portfolios. However, the function f^i is not common as it is supposed to adjust to the particular feature of each portfolio. That is, the form as well as the parameters may depend on the sample size and particularly across ages. The form of the latter will be common and defined to be of polynomial form. However, the degree will be adapted to each portfolio.*

The specification in Equation (1) is a simple variation of Cox's proportional hazards regression model. This was considered, for example, in Anderson and Senthilselvan (1980) and Gray (1990) using a known link function but with covariates that adjust the mortality given the observed heterogeneity. Cox's general

https://doi.org/10.1017/asb.2018.4 Published online by Cambridge University Press

model, in the presence of covariates, with an unknown link function is considered in Wang (2001, 2004) who suggested a local likelihood approach to estimate the function f^i. Formally, under the aforementioned assumptions, the likelihood functional $\mathcal{L}(\varphi^i; \beta^i)$ in the presence of left truncation and right censoring is given as follows:

$$\mathcal{L}(\varphi^i; \beta^i) = \prod_{e \mid Y_e^i \leq Z_e^i} \left[\left(\varphi^i_{Z_e^i}\right)^{\Delta_e^i} \exp\left(\int_{Y_e^i}^{Z_e^i} \varphi_s^i \mathrm{d}s \right) \right].$$

Therefore,

$$\begin{aligned} \log \mathcal{L}(\varphi^i; \beta^i) &= \sum_{e \mid Y_i^k \leq Z_i^k} \left[\Delta_i^k \log\left(\varphi^k_{Z_i^k}\right) - \int \mathbf{1}_{\{Y_i^k \leq s < Z_i^k\}} \varphi_s^k \mathrm{d}s \right] \\ &= \int \log(\varphi_s^k) \mathrm{d}N^k(s) - L^k(s) \varphi_s^k \mathrm{d}s, \end{aligned}$$

where $N^i(x) = \sum_{e=1}^{I_i} N_e^i(x)$ and $L^i(x) = \sum_{e=1}^{I_i} L_e^i(x)$. Given the above, we consider the local likelihood model which fits a polynomial model locally within a smoothing window. To this end, the localized log-likelihood at an age x can be written as follows:

$$\log \mathcal{L}^{\mathrm{loc}}(\varphi_x^i; \beta^i) = \int \omega_h(s-x) \log(\varphi_s^i) \mathrm{d}D^i(s) - \omega_h(s-x) L^i(s) \varphi_s^i \mathrm{d}s, \tag{2}$$

where $\omega_h(u)$ is a weight function with a bandwidth parameter $h > 0$ that assigns larger weights to observations close to x. These considerations will yield the local kernel weighted log-likelihood estimation of the polynomial function f^i. Such a formulation complies with the literature on local polynomial hazard estimation, see Fan and Gijbels (1995), Jiang and Doksum (2003) and Gámiz *et al.* (2016). We assume that $f^i(x_j)$ is a pth degree polynomial in x_l, where x_l is an element in the neighborhood of x_j. Formally, denoting $\mathbf{x}_l = \left(1, x_l - x_j, \ldots, (x_l - x_j)^p\right)^\top$ and $\beta^i = \left(\beta_0^i, \ldots, \beta_p^i\right)^\top$, we can write $f^i(x_j)$ in the following form: $f^i(x_j) = \mathbf{x}_l^\top \beta^i$.

Remark 2. *i. The reference mortality φ_x^{ref} is constructed using the aggregate data stemming from the portfolios, which will underpin the use of the common baseline. However, it may be of interest to consider a full Cox model taking into account the specific features of each portfolio. This is, for instance, investigated in Nielsen and Sandqvist (2005), where it is taken into account that mortality rates should not be around a common mean, but around a Cox regression instead. By doing so, it allows the approach to be used even when the lines of mortality are different, as long as they fit into a proportional hazard framework, see also Gustafsson et al. (2006, 2009).*

ii. Various forms of the function f^i have been considered in empirical actuarial science. For example, in Currie (2016), the function f^i has the parametric form $f^i(x) = \beta_0^i + \beta_1^i x$ for some unknown parameters β_0^i and β_1^i. Other examples were considered in Renshaw et al. (1996).

2.2. Local likelihood smoothing of mortality in discrete time

Up to now, we have considered the lifetimes of individuals to be continuous random variables subject to random censorships. In our case, the mortality data at our disposal are divided into discrete annual numbers of death occurrences and exposures. Therefore, these data only allow the application of an approach based on an approximation of the continuous filtered model in Equation (2). As noted before, both the continuous and the discrete formulations have been intensively explored in the literature, see e.g. Fan and Gijbels (1995), Jiang and Doksum (2003) and more recently, Gámiz *et al.* (2016). The latter provides a theoretic treatment of local linear mortalities and it also describes in detail the relationship between discrete and continuous sampling. In actuarial literature, early works based on discrete data date a long way back to Gram (1879, 1883) who develops local polynomial hazard estimators that are not far in spirit from our work.

The discretization of Equation (2) relies on an aggregation of the lifetimes into intervals. In this subsection, we describe a modification of the local linear estimator for discrete data in Equation (2). We suppose that the following yearly aggregated values of occurrences and exposures are available:

$$D_{x_j}^i = \sum_{e=1}^{I_i} \int_{x_j}^{x_{j+1}} \mathrm{d}N_e^i(s), \quad E_{x_j}^i = \sum_{e=1}^{I_i} \int_{x_j}^{x_{j+1}} L_e^i(s)\mathrm{d}s. \tag{3}$$

These refer, respectively, to the number of deaths and the number of individuals who are at risk in the age interval $[x_j, x_{j+1}[$. Moreover, we assume a piecewise constant hazard rate φ_x^i in the sense that $\varphi_x^i = \varphi_{x_j}^i$ for any $x \in [x_j, x_{j+1}[$. Then, a natural approximation of the localized likelihood function in a neighborhood of x_j, i.e. Equation (2), would be

$$\begin{aligned}\log \mathcal{L}(\varphi_{x_j}^i; \beta^i) &= \sum_{l=1}^{m} \omega_h(x_l - x_j)\log(\varphi_{x_l}^i) D_{x_l}^i - \omega_h(x_l - x_j)\varphi_{x_l}^k E_{x_l}^i \\ &= \sum_{l=1}^{m} \omega_{lj}\mathbf{x}_l^\top \beta^i D_{x_l}^i - \omega_{lj}\varphi^{\mathrm{ref}} e^{\mathbf{x}_j^\top \beta^i} E_{x_l}^i + C^i, \end{aligned} \tag{4}$$

where C^i is a constant offset, which does not depend on the parameter vector β^i.

Remark 3. *The true likelihood given in Equation (4) can be recovered, up to a constant offset, using the hypothesis of Poisson distributed death occurrences. In fact, if the parameter of the Poisson distribution is assumed to be* $E_x^i \varphi_x^i$ *where the intensity* φ_x^i *is as in Equation (1), then one can write the problem as a generalized linear model such that the first moment of* D_t^i *can be written as follows:*

$$\log \mathbb{E}[D_x^i] = \log E_x^i + \log \varphi_x^i = \log E_x^i + \log \varphi_x^{ref} + f^i(x),$$

where the term $\log E_x^i$ *is an offset. Then, in the presence of unknown link function* f^i*, we can rely on a localized likelihood version which adds a weight to the observations at each age. Such an approach was used to graduate life tables with attained age context in Delwarde et al. (2004), Debón et al. (2006) and Tomas (2011).*

In Equation (4), the non-negative weights, i.e. ω_{lj}, depend on the distance between the observations and the fitting point x_j and can be characterized using the kernel ω_h as follows:

$$\omega_{lj} = \begin{cases} \omega(|x_l - x_j|/h), & \text{if } |x_l - x_j| \leq h, \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where ω is Gaussian kernel and h is a smoothing parameter determining the radius of the neighborhood of x_j used in the smoothing. It gives the bandwidth of the neighborhood used in the kernel. For instance, the smaller h is the thiner the neighborhood that contributes to the likelihood at each attained age is.

In order to estimate the parameters vector β^i, we maximize the log-likelihood in (4). To this end, we let $\mathbf{D}^i = (D_{x_1}^i, \ldots, D_{x_m}^i)$ and $\varphi^i = (\varphi_{x_1}^i, \ldots, \varphi_{x_m}^i)$. Then, taking the derivative, with respect to β^i, yields the following system of equations:

$$(\mathbf{X}^j)^\top \mathbf{W}^j (\mathbf{D}^i - \varphi^i) = 0, \tag{6}$$

where $\mathbf{X}^j$ is the $m \times (p+1)$ matrix

$$\mathbf{X}^j = \begin{pmatrix} 1 & x_1 - x_j & (x_1 - x_j)^2 & \cdots & (x_1 - x_j)^p \\ 1 & x_2 - x_j & (x_2 - x_j)^2 & \cdots & (x_2 - x_j)^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_m - x_j & (x_m - x_j)^2 & \cdots & (x_m - x_j)^p \end{pmatrix}, \tag{7}$$

and $\mathbf{W}^j$ is the $m \times m$ diagonal weight matrix with diagonal elements w_{lj}, for $l = 1, \ldots, m$. Since φ^i is non-linear on β^i, the solution of the above equation, i.e. estimations, must be obtained numerically using, for example, an iterative algorithm like Nelder–Wedderburn, Newton–Raphson algorithms or the Fisher scoring methodology, see Loader (2006, Chapter 12) for further development. From these, we can get the estimation of β^i and φ^i denoted, henceforth, by $\widehat{\beta^i}$ and $\widehat{\varphi}^i$.

https://doi.org/10.1017/asb.2018.4 Published online by Cambridge University Press

2.3. Inference of the graduated mortality

The aim of this subsection is to characterize the statistical feature of the estimators considered above. We recall some well-known results in the literature on non-parametric smoothing, see e.g. Tibshirani and Hastie (1987) and Wand and Jones (1994), particularly regarding the variance of the graduated mortality and the expected behavior of these estimations. In fact, using theoretical results concerning bias and variance, the estimator $\widehat{\varphi}^i$ is shown to be asymptotically robust and consistent. It is, for instance, shown in Fan and Gijbels (1996) that the smoothed mortality rates $\widehat{\varphi}^i$ are unbiased estimators of φ^i in the sense that

$$\mathbb{E}[\widehat{\varphi}^i] \approx \varphi^i. \tag{8}$$

This approximation is found by studying the mean squared errors, which are commonly used to assess the bias of the estimation in such a framework. Expressions of the latter are available in classic textbooks and the readers can refer to Fan and Gijbels (1996) who provide an approximation to the bias of the estimator $\widehat{\varphi}^i$. Unlike the linear model fitting, there is no exact expression for moments of $\widehat{\varphi}^i_x$ due to the non-linearity in Equation (6). Using a multivariate version of Taylor series expansion around $\boldsymbol{\beta}^i$ allows us to use well-known results on the inference of the estimated parameter $\widehat{\boldsymbol{\beta}}^i$. Note that this approximation depends on the bandwidth of the neighborhood h used in the kernel. More precisely, the bias decreases with the bandwidth. This is particularly reasonable in practice, because a large bandwidth induces a miss-fitting of the local polynomials and hence also the sum of squared residuals. Furthermore, to derive the second-order moment of $\widehat{\boldsymbol{\varphi}}^i$, a variance approximation based on Taylor linearization is also generally suggested and shown to be consistent, see Loader (2006). More precisely, we have the following expression for the variance:

$$\mathbb{V}\text{ar}(\widehat{\varphi}^i) = (\mathbf{S}^i\mathbf{S}^{i\top})\widehat{\varphi}^i, \tag{9}$$

where the matrix $\mathbf{S}^i$ is given as follows:

$$\mathbf{S}^i = \begin{pmatrix} s_1^i(x_1) & s_2^i(x_1) & s_3^i(x_1) & \cdots & s_n^i(x_1) \\ s_1^i(x_2) & s_2^i(x_2) & s_3^i(x_2) & \cdots & s_n^i(x_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_1^i(x_n) & s_2^i(x_n) & s_3^i(x_n) & \cdots & s_n^i(x_n) \end{pmatrix}, \tag{10}$$

with rows $\mathbf{s}^i(x_j)^\top = (s_1^i(x_j), s_2^i(x_j), \ldots, s_n^i(x_j)) = (\mathbf{X}^{j\top}\mathbf{W}^j\mathbf{X}^j)^{-1}\mathbf{X}^{j\top}\mathbf{W}^j$, where $\mathbf{W}^j$ is the weight matrix and $\mathbf{X}^j$ is given in Equation (7).

2.4. Small-sized portfolios and sampling bias

It is worth mentioning that the relational (proportional) model considered in Equation (1) implicitly accounts for differential mortality that may arise due to portfolio specific features, e.g. particular socioeconomic groups involved, income level, etc. This is all the more true if we consider the national mortality as a baseline considering that insured portfolios show a typical behavior compared to a national mortality. Specifically, the mortality of an insured population is significantly lower than that of the national population from which it is drawn. On the other hand, when it comes to studying the mortality at a single portfolio level, some stylized facts arise which might compromise the efficiency of the graduation procedure. For instance, insured populations are generally of small size, so no or very few deaths are observable at certain ages. Therefore, the use of model (1) and the local-likelihood-based estimation procedure advocates using the information stemming from the adjacent ages to construct the mortality curve. This *learning* procedure will enhance the determination of the mortality at a given age. However, when successive ages lack information, the approach exposed above will need a large bandwidth h for the estimator to access distant ages with sufficient and reliable information. By doing so, we increase the bias surrounding the smoothed curve. Indeed, as noted before, the mean of squared errors measuring the bias due to the local regression increases with the bandwidth h.

Due to these different sources of uncertainty, we suppose that the *true* mortality curve φ_x^i, for $x = x_1 \ldots, x_n$, is known up to an unobservable multiplicative factor Θ_x^i. In other words, the portfolios examined should be regarded as a sample of the reference. Estimates based on the data will be subject to sampling errors and the smaller the group is, the bigger the relative random errors in the number of deaths will be and the less reliable the resulting estimates will be. This argument is extended to include the bias stemming from the attained age level due the consideration exposed above. Thus, if one has estimated the curve using the non-parametric approach, the *true* curve is an adjustment of the latter as multiplied by the random and non-observable parameter Θ_x^i. Such a setting is inspired by the credibility approach to hazard estimation of Nielsen and Sandqvist (2000).

3. Company-specific relative risk level

Recall that we have at our disposal K portfolios with individuals' ages ranging from x_1 to x_{n_i}. Here, the n_i's could all be different in order to be in line with insurance practices. This kind of information structure is similar to the so-called *unbalanced* framework used in actuarial science. For the sake of readability, without loss of generality, we will henceforth assume similar observed age groups for all companies, i.e. $n = n_1 = \cdots = n_K$. With a slight variation to the model, however, it can be easily extended to the unbalanced case.

3.1. The credibility model

Given the specific parameterization of the problem, one may consider the K portfolios to be subsets of the reference population and where each portfolio is characterized by a risk profile. The latter is due to the heterogeneous sizes of the portfolios as well as the underlying guarantees (for the same underlying risk). These sources of heterogeneity might also induce an age-varying risk profile within the same portfolio. Therefore, for a company i, we let the vector $\Theta^i = \text{diag}(\Theta^i_{x_1}, \ldots, \Theta^i_{x_n})$ be its relative risk level. For $x \in \{x_1, \ldots, x_n\}$, each Θ^i_x characterizes the age-specific risk level, and is an unobservable random variable.

The primary objective is to characterize the force of mortality of each company i at a specific age x through the proportional relationship introduced in Section 2.2, i.e.

$$\varphi^i = \Theta^i \alpha, \tag{11}$$

where $\alpha = (\alpha_{x_1}, \ldots, \alpha_{x_n})^\top$ such that for $j = 1, \ldots, n$, we have $\alpha_{x_j} = \exp[f^i(x_j)]\varphi^{\text{ref}}_{x_j}$. This model suggests that for each company i, the age-specific experienced force of mortality varies around the baseline α_x, which can be seen as a reference or best-estimate mortality. This fluctuation is modeled by a heterogeneity parameter Θ^i_x capturing the individual properties (heterogeneity) of each company at attained age x. Thus, using new incoming data should allow the updating of the next-period mortality φ^i_x by adjustment following the model in Equation (11). The approach is to first find an estimator $\widehat{\varphi}^i_x$ of $\varphi^i_x = \Theta^i_x \alpha_x$ for each company i using the likelihood-based approach introduced in Section 2. Henceforth, the notation $\widehat{\varphi}^i_x|\Theta^i_x$ refers to the estimation of the quantity $\alpha_x = \exp[f^i(x)]\varphi^{\text{ref}}_x$, which, by abuse of language, refers to the estimated mortality conditional to the risk profile.

Remark 4. *This reasoning is built upon the work of Nielsen and Sandqvist (2000, 2005) and Gustafsson et al. (2006). The former considered hazard rates of different groups assuming that the hazard of each group fluctuates across a common baseline hazard and used continuous sampling of observations as in Section 2.1. In the current work, we consider a discretization of the model in Nielsen and Sandqvist (2000) as the mortality data that we will use are divided into discrete annual numbers of death occurrences and exposures. Moreover, we slightly extend this framework by considering a multivariate setting and allow for the age to influence the estimation of future mortality. Using a multivariate framework will provide a base to catch the sample bias properties at the attained age level. This is even more significant given that the mortality intensities are correlated not only at the portfolio level but also between different portfolios.*

The random variables $\widehat{\varphi}^i_{x_1}, \ldots, \widehat{\varphi}^i_{x_n}$ are assumed to be dependent. Namely, the force of mortality of one age does directly impact those of other ages. This is mainly due to the graduation of mortality at a given age, which weights the information stemming from the adjacent age groups, see Section 2. This induces a dependency which will be explored later in this section. Finally, in order to characterize the next-period mortality level, we make use of credibility theory. For this purpose and using the usual credibility setting, we shall make the following assumptions:

https://doi.org/10.1017/asb.2018.4 Published online by Cambridge University Press

A1. The random vectors Θ^i are independent across companies and ages. Moreover, for $i = 1, \ldots, K$, Θ^i_x's are identically distributed with $\mathbb{E}[\Theta^i] = \mathbf{I}_n$ and $\mathbb{V}\mathrm{ar}(\Theta^i) = \boldsymbol{\sigma}$, where $\boldsymbol{\sigma}$ is a diagonal matrix with elements σ_x and $\mathbf{I}_n$ is the identity matrix.

A2. The random vectors (φ^i, Θ^i), $i = 1, \ldots, K$ are independent across companies.

A3. $\varphi^i_{x_1}, \ldots, \varphi^i_{x_n}$ are conditionally independent given Θ^i.

The first assumption (A1) ensures that the baseline mortality produces the a priori expected number of deaths under the model assumption (1), in the sense that $\mathbb{E}[D^i_x] = \mathbb{E}[\Theta^i_x \alpha_x] = \alpha_x$. The assumption (A2) means that the risk profiles are independent across portfolios. In other words, the successive realizations of the mortality intensity (as well as for the death counts) for any portfolio are independent of each other except through the risk parameter. This assumption is in line with the empirical studies and it is commonly used in actuarial literature when dealing with mortality risk. However, note that this only makes sense for mortality-contingent contracts. Thus, we should exclude annuities and pension policies where a dependence over observations is present due, for instance, to the cohort effect. Finally, assumption (A3) translates the dependency of the mortality across ages. It is only captured by the vector Θ^i. Conditionally to the latter, the forces of mortality at the age level are independent.

As noted before, $\widehat{\varphi}^i|\Theta^i$ is the *conditional* local-likelihood estimator of the intensity in Equation (1) based on the data from the ith portfolio as developed in Section 2. In view of the assumptions (A1)–(A3), it is important to recall that conditional to the knowledge of the risk profile Θ^i, the theoretical properties of $\widehat{\varphi}^i$ are identical to those of the local-likelihood estimator considered in Section 2.3. This will be used, among others, in the following lemma, in order to state some fundamental features of the dependence structure.

Lemma 1. *Under assumptions (A1)–(A3) and the notation above, we can write that*

i. The first-order moment of φ^i is given by

$$\mathbb{E}[\varphi^i] = \alpha. \tag{12}$$

ii. The variance matrix of $\widehat{\varphi}^i|\Theta^i$, denoted $\Sigma^i(\Theta^i) = \mathrm{Var}(\widehat{\varphi}^{\mathrm{i}}|\Theta^{\mathrm{i}})$, is given by

$$\Sigma^i(\Theta^i) = (\mathbf{S}^i\mathbf{S}^{i\top})\varphi^i. \tag{13}$$

Hence, the variance $\Sigma^i = \mathrm{Var}(\widehat{\varphi}^{\mathrm{i}})$ can be written as

$$\Sigma^i = \big(\mathbf{S}^i\mathbf{S}^{i\top} + \sigma\big)\alpha. \tag{14}$$

iii. The covariance of φ^i_x with $\widehat{\varphi}^i$ is given by

$$\mathrm{Cov}\left(\varphi^{\mathrm{i}}_{\mathrm{x}}, \widehat{\varphi}^{\mathrm{i}}\right) = \left(\sigma_{\mathrm{x}}\alpha_{\mathrm{x}}\right)^2 \mathbf{e}_{\delta_{\mathrm{x}}}, \tag{15}$$

with $\delta_x = j$ *if* $x = x_j$ *and* $\mathbf{e}_j$ *is the vector with all* 0*'s except for a* 1 *in the* jth *coordinate.*

Proof. To show these results, we make an intensive use of the law of total variance.

i. Equation (12) is a direct consequence of assumption (A1) which gives $\mathbb{E}[\varphi^i|\Theta^i] = \Theta^i\alpha$.
ii. The conditional variance $\Sigma^i(\Theta^i)$ is directly derived from the calculus in Section 2.3. Hence, to check (14), the law of total variance gives

$$\begin{aligned}\Sigma^i &= \mathbb{E}\big[\mathbb{V}\mathrm{ar}(\widehat{\varphi}^i|\Theta^i)\big] + \mathbb{V}\mathrm{ar}\big(\mathbb{E}[\widehat{\varphi}^i|\Theta^i]\big),\\ &= (\mathbf{S}^i\mathbf{S}^{i\top})\mathbb{E}[\widehat{\varphi}^i] + \mathbb{V}\mathrm{ar}(\Theta^i)\alpha = \big(\mathbf{S}^i\mathbf{S}^{i\top} + \sigma\big)\alpha.\end{aligned}$$

iii. Finally, to prove (15), notice that $\mathbb{C}\mathrm{ov}(\varphi^i_x, \widehat{\varphi}^i|\Theta^i) = 0$. Thus,

$$\begin{aligned}\mathbb{C}\mathrm{ov}\big(\varphi^i_x, \widehat{\varphi}^i\big) &= \mathbb{C}\mathrm{ov}\big(\mathbb{E}[\varphi^i(x)|\Theta^i], \mathbb{E}[\widehat{\varphi}^i|\Theta^i]\big) + \mathbb{E}\big[\mathbb{C}\mathrm{ov}(\varphi^i_x, \widehat{\varphi}^i|\Theta^i)\big],\\ &= \mathbb{C}\mathrm{ov}(\Theta^i_x\alpha_x, \Theta^i\alpha) = \big(\sigma_x\alpha_x\big)^2\mathbf{e}_{\delta_x},\end{aligned}$$

where the last equality follows from the independence assumption in (A1). ■

3.2. The next-period linear per-age mortality estimator

The goal is to predict the future force of mortality for each company i at the age level. Therefore, we will be looking for the inhomogeneous credibility predictor corresponding to the linear estimators of φ^i_x. Hence, we solve the following optimization problem:

$$\min_{c^i_{0,x}, \mathbf{c}^i_x} \mathbb{E}\left[\left(\varphi^i_x - c^i_{0,x} - \mathbf{c}^{i\top}_x\widehat{\varphi}^i\right)^2\right], \tag{16}$$

where $c^i_{0,x} \in \mathbb{R}$ and $\mathbf{c}^i_x \in \mathbb{R}^n$. This formulation suggests adjusting the next-period force of mortality at a given age using the information stemming from the other age groups. This should enhance the prediction for ages with low or sparse information using the credibility in ages of high information. Based on Proposition 1, we can easily derive the inhomogeneous credibility estimators of φ^i. Indeed, we can state the following proposition.

Proposition 1. *The point estimate of the linear factors in* (16) *can be written as follows:*

$$c^i_{0,x} = \big(\mathbf{1}_n - \mathbf{c}^i_x\big)^\top\alpha \quad \textit{and} \quad \mathbf{c}^i_x = \big(\sigma_x\alpha^i_x\big)^2(\Sigma^i)^{-1}\mathbf{e}_{\delta_x}. \tag{17}$$

The next-period predicted mortality estimator $\widetilde{\varphi}^i$ *of* φ^i *is given by*

$$\widetilde{\varphi}^i = (\mathbf{I}_n - \big(\alpha(\Sigma^i)^{-1}\sigma\alpha\big)^\top)\alpha + \big(\alpha(\Sigma^i)^{-1}\sigma\alpha\big)^\top\widehat{\varphi}^i. \tag{18}$$

Proof. Let us first derive the intercept $c^i_{0,x}$. To do this, we develop the expectation in Equation (16) and take the derivative with respect to $c^i_{0,x}$. This yields to the following equality:

$$c_{i,0} + (\mathbf{c}^i_x)^\top \mathbb{E}[\widehat{\varphi}^i] = 1.$$

Moreover, differentiating the expectation in Equation (16) with respect to vector $\mathbf{c}^i_x$ gives rise to the following variance:

$$\mathbb{V}\mathrm{ar}\left(\varphi^i_x - (\mathbf{c}^i_x)^\top \widehat{\varphi}^i\right),$$

needed to fully characterize the solution. This can be computed using results in Lemma 1. Indeed, we can write

$$\mathbb{V}\mathrm{ar}\left(\varphi^i_x - (\mathbf{c}^i_x)^\top \widehat{\varphi}^i\right) = \mathbb{V}\mathrm{ar}(\varphi^i_x) - 2(\mathbf{c}^i_x)^\top \mathbb{C}\mathrm{ov}(\varphi^i_x, \widehat{\varphi}^i) + (\mathbf{c}^i_x)\Sigma^i(\mathbf{c}^i_x)^\top.$$

Taking the derivative with respect to the vector $\mathbf{c}^i_x$ yields

$$2\mathbb{C}\mathrm{ov}(\varphi^i_x, \widehat{\varphi}^i) - 2\Sigma^i \mathbf{c}^i_x = 0.$$

The terms Σ^i and $\mathbb{C}\mathrm{ov}(\varphi^i_x, \widehat{\varphi}^i)$ are given in Lemma 1, which concludes the proof. ■

Note that we are able to estimate all the components needed to characterize the next-period intensity $\widetilde{\varphi}^i$, except for the variance σ. Remarking that $\widehat{\varphi}^i$ is an estimator of $\varphi^i = \Theta^i\alpha$, we can write $\widehat{\Theta}^i = \mathrm{diag}(\widehat{\varphi}^i \oslash \alpha)$, with "$\oslash$" being the Hadamard division (element-wise) operator. Therefore, a natural choice for the estimator of $\boldsymbol{\sigma}$ is

$$\widehat{\boldsymbol{\sigma}} = \left(\widehat{\Theta}^i - \mathbf{I}_n\right)^\top \left(\widehat{\Theta}^i - \mathbf{I}_n\right). \tag{19}$$

We can now derive the following estimation of the adjustment factor Θ^i.

Lemma 2. *The optimal credibility estimator of Θ^i is given by*

$$\widetilde{\widetilde{\Theta}}^i = (\mathbf{I}_n - \left(\alpha(\Sigma^i)^{-1}\widehat{\sigma}\alpha\right)^\top)\mathbf{1}_n + \left(\alpha(\Sigma^i)^{-1}\widehat{\sigma}\alpha\right)^\top \widehat{\Theta}^i,$$

and the next-period prediction of φ^i can be approximated by $\widetilde{\widetilde{\Theta}}^i\alpha$.

Remark 5. *The adjustment procedure described in Proposition 1 and Lemma 2 can be written for each individual age x in the classic form $\widetilde{\varphi}^i_x = (1 - z^i_x)\alpha_x + z^i_x\widehat{\varphi}^i_x$, where z^i_x is the credibility factor given as follows:*

$$z^i_x = (\alpha_x)^2\widehat{\sigma}^2_i\left[(\alpha_x)^2\widehat{\sigma}^2_i + \widehat{\varphi}^i_x\|\mathbf{s}^i(x)\|^2\right]^{-1}.$$

Here, recall that $\|\mathbf{s}^i(x)\|^2 = \sum_{j=1}^n \left(s^i_j(x)\right)^2$ and measures the reduction in variance of the smoothed mortality curve $\widehat{\varphi}^i_x$.

https://doi.org/10.1017/asb.2018.4 Published online by Cambridge University Press

Remark 6. *All the ingredients required to implement the credibility approach in Lemma 2, in order to predict the next-period estimator, are already determined. However, we still need to characterize an estimation of α. To do this, we borrow the same procedure considered in Nielsen and Sandqvist (2000), which amend to estimate α as a linear weighted average across the portfolios.*

4. Numerical analysis

4.1. Source of data

The data come from studies conducted by the *Institut des Actuaires*. These include in total 14 portfolios covering the period 2007–2011, with each company contributing data for at least four of a possible five years. Table 1 presents the observed characteristics of the male population of these portfolios. For this dataset, we consider a period of $T = 4$ years for all companies. The remaining year serves to test the predictive feature of the model using an in-sample analysis. The considered analysis follows similar lines as in Salhi *et al.* (2016), which also exploits the same dataset. Therefore, the age band for all companies ranges from $x = 30$ to 95 years old. Figure 1 shows the age distribution of the portfolios (in percentage), i.e. the aggregate number of individuals exposed to risk at each attained age. It graphically depicts the size heterogeneity observed between the portfolios with policyholders having different coverages. These portfolios are not only of different sizes but also of different age pyramids. For example, portfolio P1 corresponds to a typical death contingent coverage. In fact, the latter has a concentration on middle-aged populations with little exposure at high ages. When portfolios, such as P2, are concerned, we should note that they are not contingent to life annuities but rather correspond to death insurance coverage and saving contracts. They allow for a tax-advantaged investment component for those anticipating their succession and/or suitable for estate planning, which typically attracts the more elderly.

In the sequel, the baseline mortality φ_x^{ref} is a market table, denoted *IA2013*. The latter is derived from mortality trends originating from the INSEE table, the French national bureau of statistics, constructed for the French insurance market provided by the *Institut des Actuaires*, see Tomas (2011).

Before proceeding to the implementation of the methodology developed in the previous sections, we must look deeper into the particular features of our dataset. Specifically, we must focus on those where specific concerns may arise when it comes to the graduation of a mortality table using the smoothing procedure considered in Section 2. As previously reported, the experienced mortality does not only suffer in presence of a small sample size but also the under-representation of those within some age groups. This is typically the case of portfolio P2, see Table 1 and Figure 1. In fact, we have a small sample size of 7, 589 individuals with only 2% aged under 60. This is also the case for portfolios P7, P8 and to some extent P10, but with a larger exposure. For these portfolios,

TABLE 1

OBSERVED CHARACTERISTICS OF PORTFOLIOS' POPULATION.

	Period of Observation		Mean Age		Average	Mean age	
	Beginning	End	In	Out	Exposure	at Death	Size
P1	01/01/07	12/31/11	36.96	39.74	2.77	68.78	616,390
P2	01/01/07	12/31/11	69.3	73.35	4.05	80.34	7,589
P3	01/01/07	12/31/10	40.16	43.1	2.94	71.77	80,086
P4	01/01/07	12/31/11	37.5	41.13	3.63	54.08	93,165
P5	01/01/07	12/31/11	36.9	39.1	2.2	59.31	21,540
P6	01/01/07	12/31/10	48.5	52.11	3.62	82.34	847,469
P7	01/01/07	12/31/11	66.65	71.29	4.64	73.68	89,507
P8	01/01/07	04/13/11	67.51	71.38	3.86	80.72	78,650
P9	01/01/07	06/30/11	45.97	49.6	3.62	73.17	1,556,150
P10	01/01/07	12/31/11	62.97	67.64	4.67	79.77	132,990
P11	01/01/07	12/31/11	38.89	42	3.11	56.44	420,405
P12	01/01/07	12/31/11	37.05	39.2	2.15	57.41	904,020
P13	01/01/07	12/31/11	43.01	46.89	3.88	71.03	848,757
P14	01/01/07	12/31/11	50.12	54.16	4.04	72.37	233,488

the use of the smoothing procedure in Section 2.2 has the advantage of borrowing the information in age bands where the exposure is substantially larger. This may allow the mortality curve to fulfill some required local properties such as smoothness. In fact, enlarging the smoothing window h, giving access to far distant ages, may ensure the increasing of the mortality intensity across ages, which is not only a very much sought after behavior but also a biologically reasonable quality.

4.2. Entity-specific graduated mortality

In order to implement the local-likelihood-based graduation approach in Section 2, we need to identify the fitting variables. There are several components of the local fit that must be specified: the bandwidth h, the degree of local polynomial p and the weight function. (i) The latter is assumed be a Gaussian kernel as stipulated earlier in this paper. Other types of kernels can be investigated but this has much less effect on the bias and the variance tradeoff. As noted by Loader (2006), the kernel choice only influences the visual quality of the fitted regression curve. (ii) In addition, the bandwidth has a critical effect on the local regression fit. The simplest specification is a constant bandwidth for all ages x. This is, however, not satisfactory in our case. In fact, as mentioned throughout the paper, for ages where data is available in a sufficient amount, small bandwidths will produce a convenient fit with the desired features. In turn, when the population is poorly represented at some ages, large values of h are required. Accordingly, one might choose a different bandwidth for each fitting age $x \in \{x_1, \ldots, x_m\}$,

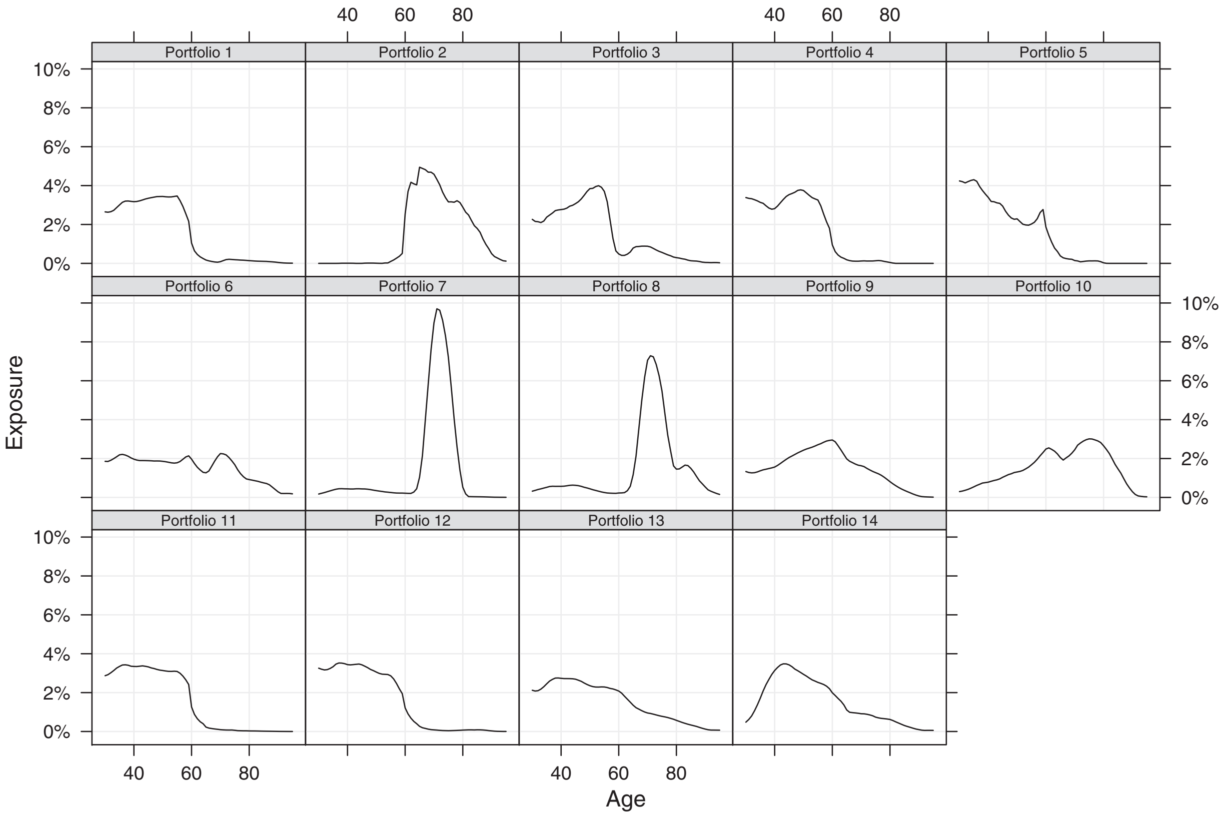

FIGURE 1: Distribution of age groups in the portfolios.

taking into account local features such as the local intensity and the amount of data. The problem of choosing the bandwidth h has received a lot of attention in the literature. See, for example, Fan and Gijbels (1995), Jones *et al.* (1996), Bagkavos and Patil (2009), Nielsen *et al.* (2009) and Gámiz *et al.* (2016) and the references therein.

First of all, when the bandwidth h does not depend on the age level, we can use a scoring procedure based on a generalization of Aïkake Information Criterion (AIC) that uses the deviance function, i.e. the likelihood together with the degrees of freedom of the fitted model, to rank the models. In our case, as we adopt a local rather than a global bandwidth, we advocate using some popular and yet efficient data-driven approaches. Here, we use the selection rule proposed by Jiang and Doksum (2003). The latter can be summarized in the following steps:

- Step 1 We choose an initial global bandwidth h. The latter can be based on a modified AIC as described above and advocated by Loader (2006). This is, for instance, the approach used in the empirical work of Tomas (2011). Then, pilot estimators $\widehat{\varphi_x}$ of φ_x are obtained by using the same bandwidth h for ages x and the local likelihood estimator in Equation (4).
- Step 2 For each age level x, we optimize the likelihood functional in Equation (4) being function of the bandwidth. We obtain its minimizer h.
- Step 3 We run a local smoother of the bandwidths h over ages using the global bandwidth in **Step 1** and the same kernel ω.

The above rule is the analog of the least-squares cross-validation or the leave-one-out principle, see Mammen *et al.* (2011), Gámiz *et al.* (2013) and Gámiz *et al.* (2016). In Gámiz *et al.* (2016), a precise connection between the cross-validation procedure and our discrete framework is investigated.

Once an estimate of the local bandwidths are obtained, one can estimate the optimal polynomial degree p through the global partial likelihood. In Table 2, we copied the degree of the polynomial used for smoothing as well as the corresponding degree of freedom and the AIC score. This is intended to represent the global sparsity of the data and the goodness of fit quality. We can see that for some portfolios the optimal choice of the degree controls induce a high level of degrees of freedom, i.e. portfolios P5 and P6. This is to say that the corresponding "smoothed" curves $\widehat{\boldsymbol{\varphi}}^i$, $i = 5, 6$, will be noisy, showing many features. Indeed, the degree of freedom is a qualitative proxy for the regularity of the graduated mortality curve as the *smoothness* evolves inversely to the degree of freedom. This feature can already be deduced from the limited amount of information (exposures) that are at our disposal for these portfolios. However, the sparsity of the data is not only represented by the exposure. Indeed, the deaths are of paramount importance when characterizing the survival rate. In fact, looking at the exposure reported in Table 1, one could expect a high degree of freedom for portfolio P2, having only 7, 589 individuals exposed to risk. However, the death records are concentrated on a small band making the

TABLE 2

LOCAL-LIKELIHOOD SMOOTHING PARAMETERS' OPTIMAL CHOICE.

	P1	P2	P3	P4	P5	P6	P7
AIC	57.12	49.34	79.68	78.91	61.10	106.73	68.37
Degree	2	2	1	1	3	3	1
DF	5	3	4	6	16	10	8
	P8	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**
AIC	74.11	63.96	53.44	70.86	74.82	82.70	78.37
Degree	2	1	2	1	1	3	2
DF	4	5	4	6	6	7	5

smoothing less noisy as the information needed to estimate the mortality at each individual age is accessible at the immediate adjacent ages. This is, for instance, not the case for portfolio P5 having more exposed individuals but with higher sparsity and a smaller number of deaths over few ages. For the remaining portfolios, the degrees of freedom are relatively small. In the following, we will implement the credibility approach described in the last section to assess the impact of the latter on the graduated mortality curves.

4.3. Next-period mortality rate

Here, we consider the mortality experience over the period 2007–2010 upon which we calibrate the smoothing procedure considered in the above subsection. For each portfolio, we build a graduated mortality table $\widehat{\varphi}^i$ and aim to adjust the latter for the next-period projection. For each age x, the graduated mortality gives a candidate rate for the next period, i.e. $\widehat{\varphi}^i_x$. The insurer has the possibility of relying on this rate or adjusting it given the experience stemming from the other rates at other ages. In other words, the mortality used for the next period forecasts can be adjusted using the credibility formula in Equation (18). To do this, we estimate the different quantities needed to implement (18) as follows:

i. Following Remark 6, the expected mortality rate α can be estimated as follows:

$$\widehat{\alpha} = \Big(\sum_{x=x_1}^{x_n} \sum_{i=1}^{K} E^i_x \widehat{\varphi}^i_x \Big) / \sum_{i=1}^{K} E^i_{\bullet}.$$

ii. The weight loading matrix S^i is given as an output of the graduation step and can be estimated using Equation (10).

https://doi.org/10.1017/asb.2018.4 Published online by Cambridge University Press

iii. The diagonal matrix σ relies on the variance of Θ^i_x's which can be estimated given that $\widehat{\Theta}^i_x = \widehat{\varphi}^i_x/\widehat{\alpha}$, and thus we can write

$$\sigma_x = \sum_{i=1}^{K} \left(\widehat{\varphi}^i_x/\widehat{\alpha}\right)^2/K - \left(\left(\sum_{i=1}^{K} \widehat{\varphi}^i_x/\widehat{\alpha}\right)/K\right)^2.$$

Figure 2 depicts, respectively, the graduated mortality over the period 2007–2010 as described above and the next period (2011) mortality rates using the credibility formula in Equation (18). Similarly, Figure 3 represents the next period predicted deaths using the two mortality rates. In these figures, we grayed areas (ages) where the relative difference between the smoothed mortality and its adjusted counterpart exceeds a 10% level. More precisely, this corresponds to the ages x where $|\widetilde{\varphi}^i_x - \widehat{\varphi}^i_x|/\widehat{\varphi}^i_x > 0.1$. At first glance, we remark that the credibility adjustment does change the mortality rate and does overall propose a smoother curve compared to the initial one, and this is even evident when dealing with portfolios with small sizes and high degrees of freedom. In fact, when we deal with portfolios such as P5, where the exposure-to-risk as well as the underlying deaths are very limited, the smoothing approach fails to capture the mortality structure and the output of the procedure proposed in Section 2 is very irregular and noisy. Indeed, as noted above, such a procedure needs information stemming from adjacent ages when a particular age lacks sufficient exposure. Here, the case of P5 provides an explicit example of the limit of the semi-parametric smoothing techniques as the limitation on the information is shared across ages. This is why the corresponding degree of freedom is high and the AIC is low, see Table 2, and explains the irregular curve (dashed line) for the smoothed mortality.

Furthermore, the degrees of freedom given as $\mathrm{tr}(\mathbf{S}^i\mathbf{S}^{i\top})$ provide information on the credibility of the smoothed curve $\widehat{\boldsymbol{\varphi}}^i$. In fact, as we can see in Equation (18) or in a more tractable way as in Remark 5, the higher the degrees of freedom, i.e. $\mathrm{tr}(\mathbf{S}^i\mathbf{S}^{i\top}) = \sum_{x=x_1}^{x_n} \|\mathbf{s}^i(x)\|^2$, the smaller the weight attached to the smoothed curve $\boldsymbol{\varphi}^i$ (in aggregate). At an age level $x \in \{x_1, \ldots, x_m\}$, one should look at the individual variance $\|\mathbf{s}^i(x)\|^2$. That being said, we can conclude that the parameter driving the adjustment at the age levels vanishes, meaning that the adjusted mortality rate $\widetilde{\varphi}^i_x$ is close to the reference α_x. It comes as no surprise, then, to find that the adjusted curve tends to offset this undesired effect thanks also to the information coming from other ages but from different portfolios.

The visual inspection of the credibility-based mortality curve shows that the regularity is preserved avoiding the limitation of the sole smoothing procedure discussed above. For some portfolios, such as portfolio P12, the regularization based on the credibility attached to each age level enhances the prediction of the future mortality. Indeed, the smoothed mortality based on past observations suggests a local distortion of the curve for ages ranging from 60 to 80. This particular feature is, however, not observed in the mortality curve for the year 2011 and thus the credibility-based curve has a better fitting. This can also

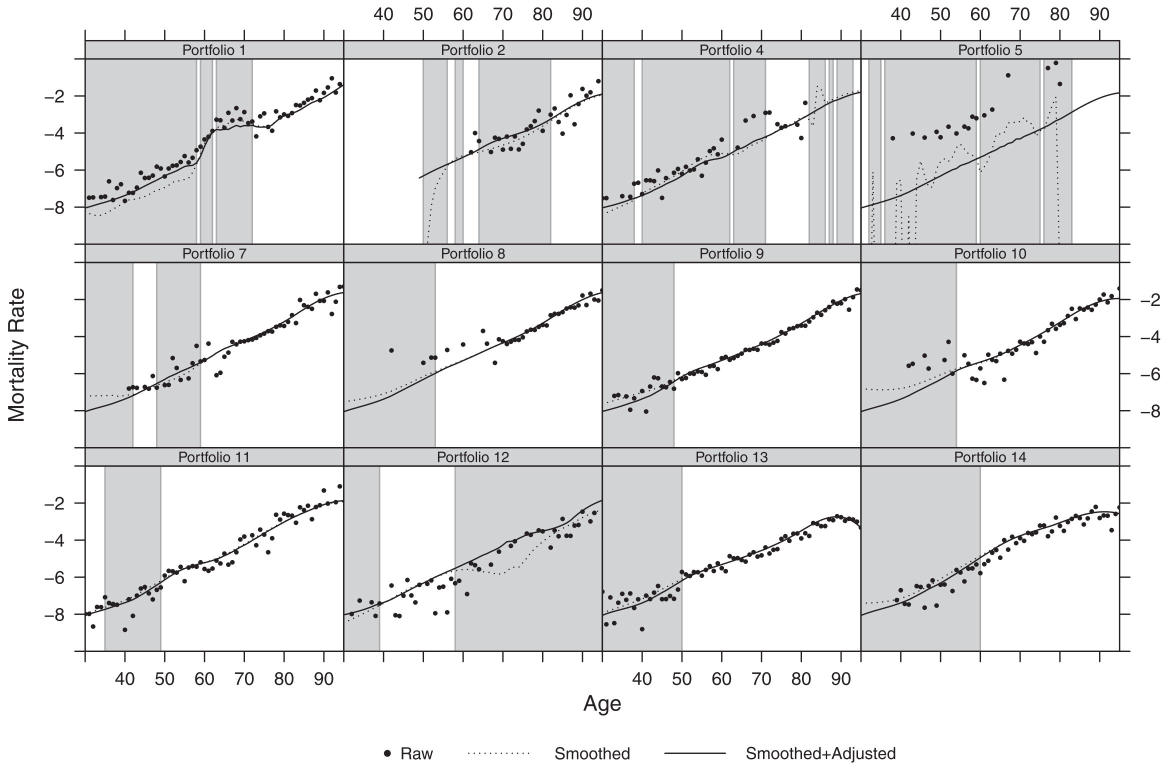

FIGURE 2: The next-period mortality rate based on the graduation of mortality (dashed line) as well as its credibility based adjustment (solid line) based on the period 2007–2010. Both predictions are compared to the observed mortality rate over the period 2011 (black dots).

https://doi.org/10.1017/asb.2018.4 Published online by Cambridge University Press

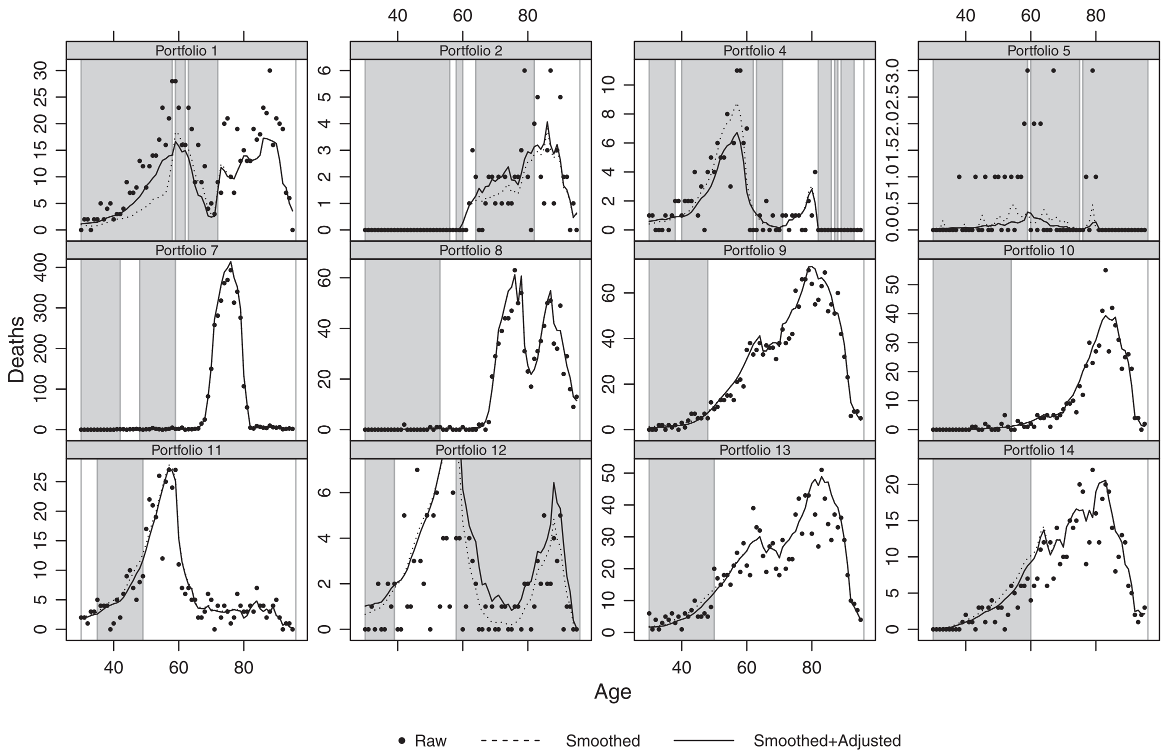

FIGURE 3: The next-period deaths prediction based on the graduation of mortality (dashed line) as well as its credibility-based adjustment (solid line) based on the period 2007–2010. Both predictions are compared to the observed deaths over the year 2011 (black dots).

be observed in Figure 3, where the predicted deaths using $\widetilde{\boldsymbol{\varphi}}^i$, for $i = 12$, are (visually) more in line with the observations. The same conclusions, in the grayed area, can be drawn for the other portfolios.

4.4. Proximity between the observations and the model

Besides the visual inspection of the proposed adjustments and in order to understand the impact of the latter, we will use some known statistics to quantify the proximity between the observations and the outputs of the two curves considered in Figures 2 and 3. We assess the overall deviation with the observed mortality by comparing criteria measuring the distance between the observations and the models with the χ^2, i.e. Forfar *et al.* (1988b), the mean average percentage error (MAPE), see for instance, Felipe *et al.* (2002), as well as the standardized mortality ratio (SMR), i.e. the ratio of deaths observed to those predicted. The quantities summarizing the proximity between the observations and the model, for each portfolio i at calendar year $t = 2011$, are described as follows:

i. The χ_i^2 allows us to measure the quality of the fit of the model. It writes

$$\chi_i^2 = \sum_{x=x_1}^{x_n} \frac{\left(D_x^i - E_x^i\,\widehat{q}_x^i\right)^2}{E_x^i\,\widehat{q}_x^i\left(1 - \widehat{q}_x^i\right)}.$$

ii. The MAPE is the average of the absolute values of the deviations from the observations

$$\mathrm{MAPE}^i = \frac{\sum_{x=x_1}^{x_n} \left|\left(D_x^i/E_x^i - \widehat{q}_x^i\right)/\left(D_x^i/E_x^i\right)\right|}{\sum_{x=x_1}^{x_n} D_x^i} \times 100.$$

iii. The SMR is computed as the ratio between the observed and fitted number of deaths in each portfolio

$$\mathrm{SMR}^i = \frac{\sum_{x=x_1}^{x_n} D_x^i}{\sum_{x=x_1}^{x_n} E_x^i\,\widehat{q}_x^i}.$$

Hence, if $\mathrm{SMR}^i > 1$, the fitted deaths are under-estimated and vice-versa if SMR < 1. Note that we can consider the SMR_i as a global criterion which does not take the age structure into account, compared to the χ_i^2 and the MAPE_i, for instance.

Table 3 summarizes the above-mentioned quantities giving the overall deviation between the observations and the adjustment analysis for portfolios P1 to P14 (except 3 and 6 which do not contain observations for year 2011) obtained by the smoothing approach together with the credibility adjustment procedure. When looking at criteria and quantities which take the age structure of the error into account, the credibility approach has an important benefit compared to the sole graduated curve. The quality of the fit increases, sometimes drastically, i.e.

TABLE 3

TESTS AND QUANTITIES SUMMARIZING THE DEVIATION BETWEEN THE OBSERVATIONS AND THE MODEL.

		Hardy–Panjer	Poisson–Gamma	Makeham–Credibility	Smoothed	Smoothed+Adj.
χ^2	Portfolio 1	1901.240	1928.680	259.400	357.870	193.967
MAPE (%)		102.660	102.000	32.870	3.018	2.349
SMR		1.737	1.756	1.126	1.487	1.385
χ^2	Portfolio 2	34.890	33.640	30.940	37.612	31.166
MAPE (%)		48.030	49.120	53.990	20.119	20.842
SMR		1.037	1.002	0.905	1.102	0.948
χ^2	Portfolio 4	130.120	132.890	79.321	58.615	51.515
MAPE (%)		95.390	92.490	44.880	14.006	13.078
SMR		0.826	0.853	1.405	0.984	1.168
χ^2	Portfolio 5	473.680	573.940	348.180	NA	370.401
MAPE (%)		85.660	88.040	90.420	59.296	56.038
SMR		2.857	3.424	5.021	3.513	5.534
χ^2	Portfolio 7	221.640	223.560	195.000	77.997	72.795
MAPE (%)		135.390	135.710	37.250	0.534	0.509
SMR		0.846	0.844	0.823	0.922	0.922
χ^2	Portfolio 8	2575.630	2583.900	2414.250	66.033	61.174
MAPE (%)		323.780	324.610	263.210	1.100	1.122
SMR		0.232	0.231	0.243	0.928	0.930
χ^2	Portfolio 9	1572.530	1573.970	1502.870	57.461	53.735
MAPE (%)		368.080	368.290	125.640	0.764	0.755
SMR		0.423	0.423	0.419	0.932	0.940
χ^2	Portfolio 10	115.820	116.470	97.880	83.790	72.448
MAPE (%)		89.680	91.030	46.140	3.356	3.530
SMR		0.871	0.862	0.960	0.948	0.950
χ^2	Portfolio 11	415.320	417.530	76.480	55.888	55.127
MAPE (%)		152.870	151.690	46.970	5.934	5.548
SMR		0.829	0.837	1.018	0.918	0.964
χ^2	Portfolio 12	130.050	129.230	90.740	88.836	76.459
MAPE (%)		110.540	107.220	95.270	36.577	33.344
SMR		0.598	0.619	0.543	0.669	0.539
χ^2	Portfolio 13	351.560	351.360	263.550	94.570	89.428
MAPE (%)		180.910	180.610	54.620	1.765	1.608
SMR		0.839	0.840	0.832	0.914	0.930
χ^2	Portfolio 14	227.860	227.950	85.920	59.317	50.885
MAPE (%)		159.740	160.600	53.530	5.659	4.852
SMR		0.792	0.788	0.939	0.827	0.860

portfolio P1, in terms of having the minimum χ_i^2 and MAPE_i values, i.e. the last panels of Table 3. Also, the credibility adjustment exhibits the highest p-value for the likelihood ratio test. Even when we consider a global indicator of the quality of the fit such as the SMR_i which does not take the age structure into account, the proposed procedure seems to perform better than the graduated curve. However, notice that the impact of adjustment is smaller when the portfolios are quite big. This is already noticeable when checking visually as mentioned earlier.

4.5. Comparaison with traditional approaches

Here, we wish to compare our model to the Hardy and Panjer (1998) and Poisson–Gamma credibility analysis applied to our mortality datasets. Moreover, we intend to compare our results to a similar approach introduced in Salhi *et al.* (2016), where the graduation of mortality is based on a parametric method, i.e. Makeham law, and the credibility theory is used to adjust the latter with incoming new data experience. The first two approaches focus on the actual to expected mortality ratio, in aggregate level, as a key observation. The adjustment is directly applied to this quantity. Specifically, the a priori expected number of deaths for each portfolio is updated at each period given the credibility weight on the observations coming from this portfolio and the one computed on the basis of the others. Table 3 also presents the tests and quantities summarizing the overall deviation between the observations for these different approaches in comparison to the one exposed in this paper.

We first note that the Hardy–Panjer and Poisson–Gamma approaches produce relatively similar graduations as the tests suggest sensibly similar outputs. However, we notice some differences with the Makeham credibility model which displays more favorable results. This is already outlined in Salhi *et al.* (2016) and this may be explained by the age-specific adjustment but also thanks to the structural feature added by the Makeham parametric model. Taking into account the age, and thus the structure of the portfolio, increases the goodness of fit of as well as the predictive performance of the constructed mortality. Regarding the local likelihood approach, we notice that the force of mortality adjusts to more complex mortality structures and thus offers a better fit for portfolios with sparse information. However, for very small portfolios, the smoothed mortality rates fail to properly predict the next-period deaths compared to the aforementioned approaches. The credibility-based revision at the age level globally enhanced the predictive ability of the graduated mortality. Specifically, when it comes to the tests that are sensitive to the age structure, we notice that the credibility-based adjustment offers an outstanding fit as the tests are favorable compared to the Hardy–Panjer, Poisson–Gamma and the Makeham-based approaches. More importantly, the main advantage of our method over these approaches is its ability to adjust to more complex mortality structures and thus offer a better fit for portfolios with sparse information, e.g. the MAPE for the following period, for Portfolio 2 (small), goes from 3.10% to 2.34%.

5. Concluding remarks

In this paper, we proposed a methodology to adjust the graduated mortality table that uses an adaptive smoothing procedure based on the local likelihood. The adjustment is based on the credibility weighting technique of the smoothed curves and a reference. Our approach takes into account the age-specific heterogeneity that may arise in real-world datasets. Therefore, we consider updating the mortality for each age based on the upcoming past information from the same age but also the neighboring ages. The inclusion of the neighboring ages is crucial as the particular smoothing procedure used in this paper adds a dependency between the single ages. Based on classic results on the inference of the smoothing procedure, we derived the closed-form formulas needed to adjust the mortality.

The proposed methodology is shown to outperform the classic credibility approaches that do not take into account the age structure of portfolios. This is in line with the recent work in this field as mentioned by Salhi *et al.* (2016). Even when the age structure is accounted for the methodology developed in this paper has an important benefit. This is mainly due to the underlying curve built using an adaptive procedure compared to the parametric model considered in Salhi *et al.* (2016).

We should note that the proposed model can be investigated in order to mathematically quantify the errors induced in the assessment of the next-period mortality curve. This leads us to consider the uncertainty stemming from the estimation of the different variables used in the updating procedure. There are also several practical issues we do not address here such as the impact on the pricing of life insurance contracts. These are open questions that we openly acknowledge and leave for future research.

Acknowledgements

This work has been supported by the BNP Paribas Cardif Chair "Data Analytics & Models in Insurance" and the ANR project "LoLitA" (ANR-13-BS01-0011). The views expressed in this paper are the authors' own and do not necessarily reflect those endorsed by BNP Paribas Cardif. Both authors would like to thank the editor and the two anonymous referees for their careful reading, their useful comments and their suggestions which improved our previous presentation.

References

Anderson, J. and Senthilselvan, A. (1980) Smooth estimates for the hazard function. *Journal of the Royal Statistical Society. Series B (Methodological)*, **42**(3), 322–327.

Bagkavos, D. and Patil, P. (2009) Variable bandwidths for nonparametric hazard rate estimation. *Communications in Statistics–Theory and Methods*, **38**, 1055–1078.

CLEVELAND, W.S. (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association,* **74**, 829–836.

COPAS, J. and HABERMAN, S. (1983) Non-parametric graduation using kernel methods. *Journal of the Institute of Actuaries,* **110**, 135–156.

COX, D. (1972) Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological),* **34**, 187–220.

CURRIE, I.D. (2016) On fitting generalized linear and non-linear models of mortality. *Scandinavian Actuarial Journal*, **2016**(4), 356–383.

DEBÓN, A., MONTES, F. and SALA, R. (2006) A comparison of nonparametric methods in the graduation of mortality: Application to Data from the Valencia Region (Spain). *International statistical Review,* **74**, 215–233.

DELWARDE, A., KACHKHIDZE, D., OLIE, L. and DENUIT, M. (2004) Modèles linéaires et additifs géneralisés, maximum de vraisemblance local et méthodes relationelles en assurance sur la vie. *Bulletin Français d'Actuariat,* **6**, 77–102.

FAN, J. and GIJBELS, I. (1995) Data-driven bandwidth selection in local polynomial fitting: Variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**(2), 371–394.

FAN, J. and GIJBELS, I. (1996) *Local Polynomial Modelling and Its Applications*, volume 66,Monographs on Statistics and Applied Probability. London: Chapman and Hall.

FELIPE, A., GUILL'EN, M. and PÉREZ-MARÍN, A. (2002) Recent mortality trends in the Spanish population. *British Actuarial Journal,* **8**, 757–786.

FORFAR, D., MCCUTCHEON, J. and WILKIE, A. (1988a) On graduation by mathematical formula. *Journal of the Institute of Actuaries,* **115** , 1–149.

FORFAR, D., MCCUTCHEON, J. and WILKIE, A. (1988b), On graduation by mathematical formula. *Journal of the Institute of Actuaries,* **115**, 1–459.

GÁMIZ, M.L., JANYS, L., MIRANDA, M.D.M. and NIELSEN, J.P. (2013) Bandwidth selection in marker dependent kernel hazard estimation. *Computational Statistics & Data Analysis,* **68**, 155–169.

GÁMIZ, M.L., MAMMEN, E., MIRANDA, M.D.M. and NIELSEN, J.P. (2016) Double one-sided cross-validation of local linear hazards. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* **78**, 755–779.

GRAM, J. (1879) *Om Rækkeudviklinger, bestemte ved Hjælp af de mindste Kvadraters Methode*. Ph.D. thesis. Copenhagen: Andr. Fred. Høst & Son.

GRAM, J. (1883) Ueber die Entwickelung reeller Functionen in Reihen mittelst der Methode der kleinsten Quadrate. *Journal für die reine und angewandte Mathematik,* **94**, 41–73.

GRAY, R.J. (1990) Some diagnostic methods for Cox regression models through hazard smoothing. *Biometrics*, **46**(1), 93–102.

GUSTAFSSON, J., NIELSEN, J., PRITCHARD, P. and ROBERTS, D. (2006) Quantifying operational risk guided by kernel smoothing and continuous credibility. *The ICFAI Journal of Financial Risk Management,* **3**, 23–47.

GUSTAFSSON, J., NIELSEN, J., PRITCHARD, P. and ROBERTS, D. (2009) Quantifying operational risk guided by kernel smoothing and continuous credibility: A practioner's view. *Journal of Operational Risk,* **1**, 43–55.

HARDY, M. and PANJER, H. (1998) A credibility approach to mortality risk. *ASTIN Bulletin,* **28**, 269–283.

HOUGAARD, P. (1984) Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika,* **71**, 75–83.

JIANG, J. and DOKSUM, K. (2003) On local polynomial estimation of hazard rates and their derivatives under random censoring. *Lecture Notes-Monograph Series*, **42**, 463–481.

JONES, M.C., MARRON, J.C. and SHEATHER, S.J. (1996) A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association,* **91**, 401–407.

LEJEUNE, M. (1985) Estimation non-paramétrique par noyaux: Régression polynomiale mobile. *Revue de Statistique Appliquée,* **23**, 43–67.

LOADER, C. (2006) *Local Regression and Likelihood.* Berlin: Springer Science & Business Media.

MAMMEN, E., MARTÍNEZ MIRANDA, M.D., NIELSEN, J.P. and SPERLICH, S. (2011) Do-validation for kernel density estimation. *Journal of the American Statistical Association,* **106**, 651–660.

NIELSEN, J. and SANDQVIST, B. (2000) Credibility weighted hazard estimation. *Astin Bulletin,* **30**, 405–418.
NIELSEN, J. and SANDQVIST, B. (2000) Proportional hazard estimation adjusted by continuous credibility. *ASTIN BULLETIN,* **35**, 239–258.
NIELSEN, J.P., TANGGAARD, C. and JONES, M. (2009) Local linear density estimation for filtered survival data, with bias correction. *Statistics,* **43**, 167–186.
RAMLAU-HANSEN, H. (1983) The choice of a kernel function in the graduation of counting process intensities. *Scandinavian Actuarial Journal,* **1983**, 165–182.
RENSHAW, A., HABERMAN, S. and HATZOPOULOS, P. (1996) The modelling of recent mortality trends in United Kingdom male assured lives. *British Actuarial Journal,* **2**, 449–477.
SALHI, Y., THÉROND, P.E. and TOMAS, J. (2016) A credibility approach of the Makeham mortality law. *European Actuarial Journal,* **6**, 61–96.
TIBSHIRANI, R. and HASTIE, T. (1987) Local likelihood estimation. *Journal of the American Statistical Association,* **82**, 559–567.
TOMAS, J. (2011) A local likelihood approach to univariate graduation of mortality. *Bulletin Français d'Actuariat,* **11**, 105–153.
VAUPEL, J., MANTON, K.G. and STALLARD, E. (1979) The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography,* **16**, 439–454.
WAND, M.P. and JONES, M.C. (1994) *Kernel smoothing.* Boca Raton, FL: CRC Press.
WANG, W. (2001) *Proportional Hazards Regression Model with Unknown Link Function and Applications to Longitudinal Time-to-event Data*. Ph.D. thesis, University of California, Davis.
WANG, W. (2004) Proportional hazards regression models with unknown link function and time-dependent covariates. *Statistica Sinica*, **14**(3), 885–905.

YAHIA SALHI
Université de Lyon
Université de Lyon 1
ISFA, Laboratoire de Sciences Actuarielle et financière
EA2429, 50 Avenue Tony Garnier
F-69007 Lyon, France
E-mail: yahia.salhi@univ-lyon1.fr

PIERRE-E. THÉROND (Corresponding author)
Université de Lyon
Université de Lyon 1
ISFA, Laboratoire de Sciences Actuarielle et financière
EA2429, 50 Avenue Tony Garnier
F-69007 Lyon, France

Galea & Associés
25 rue de Choiseul
Paris 75002, France
E-Mail: pierre@therond.fr